

Измерение информации

Объемный подход
Формулы Хартли и Шеннона
§3

Объем текстового сообщения

В целях предотвращения возможных проблем при обработке результатов участников единого государственного экзамена (ЕГЭ), а также при расчете свидетельств о результатах ЕГЭ в 2007 году Управление контроля и оценки качества образования информирует Вас о правилах внесения информации паспортных данных обучающихся в региональную базу данных участников ЕГЭ, а также в бланк регистрации при заполнении бланка участником экзамена (для субъектов Российской Федерации, использующих базовую технологию):

1. Информация о фамилии, имени, отчестве участников вносится в русские буквы (кириллицей) в именительном падеже. При необходимости запись данных, состоящих из нескольких частей, слова необходимо разделить одним пробелом или одним дефисом, как это указано в документе, удостоверяющем личность участника. Использование подчёркивания и дефиса недопустимо.

2. Поле «Серия документа» может не заполняться в случае, если документ не имеет серии. Если серия присутствует в документе, то поле может содержать:

- только арабские цифры и дефис;
- только латинские буквы;
- только латинские буквы, дефис и арабские цифры;
- только буквы кириллицы;
- только буквы кириллицы, дефис и арабские цифры;
- только буквы кириллицы, дефис и латинские буквы, используемые для записи римских цифр: «I», «II», «X», «O», «L», «M». Вместо латинской буквы «I» возможно использовать арабскую цифру «1».

Регистр используемых букв не имеет значения.

Поле не может содержать:

- пробелов;
- одновременно и латинские, и русские буквы, за исключением случая, описанного в п.2.f;
- различные комбинации символов, описанных в различных подпунктах п.2.

3. Поле «Номер документа» может содержать только цифры.

Перечисленные правила должны выполняться сотрудниками РЦОИ (ПШОИ) при организации работ по сбору паспортных данных участников экзамена, а также при обработке бланков регистрации.

Корректность файлов с результатами обработки экзаменационных материалов, предназначенных для отправки в ФГУ «Федеральный центр тестирования», должна быть проверена программой «Проверка корректности файлов», предоставляемой через сайт технической поддержки в комплекте «FileEGE».

В связи об участниках ЕГЭ, данные которых не соответствуют правилам корректности, необходимо внести изменения.

В целях предотвращения возможных проблем при обработке результатов участников единого государственного экзамена (ЕГЭ), а также при распечатке свидетельств о результатах ЕГЭ в 2007 году Управление контроля и оценки качества образования информирует Вас о правилах внесения информации паспортных данных обучающихся в региональную базу данных участников ЕГЭ, а также в бланк регистрации при заполнении бланка участником экзамена (для субъектов Российской Федерации, использующих базовую технологию):

Объем текстового сообщения не связан с его содержанием

Сравним информационные размеры текстовых страниц

Статистика:

Характеристика	Значение
Страниц:	1
Абзацев:	18
Строк:	45
Слов:	297
Знаков:	1960
Знаков и пробелов:	2240

Всего знаков на первой
странице 2240
(размер кегля 14 пт.)

Статистика:

Характеристика	Значение
Страниц:	1
Абзацев:	1
Строк:	20
Слов:	62
Знаков:	434
Знаков и пробелов:	495

Всего знаков на второй
странице 495
(размер кегля 32 пт.)

**Естественно,
объем первой страницы
больше чем второй**



Алфавитный подход

Как зависит информационный объем текстового сообщения в рукописном виде от размера (мощности) алфавита?

Информационный объем одного символа зависит от мощности алфавита по формуле Хартли:

$$i = \log_2 N \text{ (бит),}$$

где N - количество символов в алфавите

Мощность алфавита и размер одного символа в русском языке

Это не только собственно буквы
(строчные и прописные),
но и знаки препинания, скобки, кавычки и т. п.

66 букв,
знаки **!, ?, :, ;, ., , , -, —, (,), «, »**
и пробел,
всего **79 знаков**

$$i = \log_2 79 = 6,303780748 \text{ бит}$$

Кодирование русских букв для передачи телеграфом

Русский телеграф содержит 32 клавиши
(нет прописных букв, буквы ё и знаков
препинания, есть слова «зпт», «тчк»)

$$i = \log_2 32 = 5 \text{ бит,}$$

т. е. в телеграфном коде символов
в 5 раз больше,

чем в исходном тексте

закодированном с помощью

знаковой системы,



Количество информации в сообщении,
равно количеству информации,
закодированном с помощью
которое несет один знак;
знаковой системы,
умноженному на число знаков в сообщении.
равно количеству информации,
которое несет один знак,
умноженному на число знаков в сообщении.

$$I = K \cdot i$$

$i = \log_2 N$ – вес одного символа,
K - число знаков в сообщении.

Пример

**Какова мощность алфавита,
с помощью которого записано сообщение,
содержащее 2048 символов (K),
если его объем составляет 1,25 Кб (I)?**

$$I = 1,25 \text{ Кб} = 10\ 240 \text{ бит}$$

1. определим количество бит, приходящееся на один символ:

$$i = 10\ 240 \text{ бит} : 2\ 048 = 5 \text{ бит}$$

3. определим количество символов в алфавите:

$$N = 2^i = 2^5 = 32$$

Ответ: в алфавите 32 символа



Текст в электронном виде

Для определения количества информации в тексте, набранном на клавиатуре компьютера необходимо знать, с помощью какой кодировочной таблицы кодируются символы

Стандартная таблица ASCII (аски) позволяет закодировать 256 символов, т.е. длина кода одного символа равна 8 битам.

Количество символов в сообщении равно количеству байтов в объеме информации

Задача

Пользователь компьютера, хорошо владеющий навыками ввода информации с клавиатуры, может вводить в минуту 100 знаков.

Мощность алфавита, используемого в компьютере равна 256. Какое количество информации в байтах может ввести пользователь в компьютер за 1 минуту?



Единицы измерения количества информации

Принята следующая система единиц
измерения количества информации:

$$1 \text{ б (байт)} = 8 \text{ бит}$$

$$1 \text{ Кб (килобайт)} = 2^{10} \text{ б}$$

$$1 \text{ Мб (мегабайт)} = 2^{10} \text{ Кб}$$

$$1 \text{ Гб (гигабайт)} = 2^{10} \text{ Мб}$$

Единицы измерения количества информации

1 Тб (терабайт) = 2^{10} Гб

1 Пб (петабайт) = 2^{10} Тб

1 Эб (эксабайт) = 2^{10} Пб

1 Зб (зетабайт) = 2^{10} Эб

1 Йб (йоттабайт) = 2^{10} Зб

$$2^{10} = 1024$$



Домашнее задание

§3, Вопросы и задание стр.20

1. Найти x из следующих соотношений:

а) 16^x бит = 32 Мбайт;

б) 8^x Кбайт = 16 Гбайт.

2. Решить систему уравнений:

$$\begin{cases} 2^{x+2} \text{ (бит)} = 8^{y-5} \text{ (Кбайт)} \\ 2^{2y-1} \text{ (Мбайт)} = 16^{x-3} \text{ (бит)} \end{cases}$$